Strong AI and Chinese Room Argument

Vivek Atulkar - 110050039
Vivek Agarwal – 110050042
Shivam H Prasad - 110050041
Mayank Meghwanshi - 110050012

Motivation

- Work in Artificial Intelligence (AI) has produced computer programs that can beat the world chess champion , and converse in natural language.
- Our experience shows that these activities require understanding and intelligence.
- Does computer prowess at challenging games and conversation then show that computers can understand and be intelligent?
- Will further development result in digital computers that fully match or even exceed human intelligence?

Weak AI and Strong AI

- According to **weak AI** hypothesis, machines can demonstrate intelligence, but do not necessarily have a mind, mental states or consciousness.
  - E.g. - play chess, read addresses, drive cars, recognise speech and so on.

- In contrast, **strong AI** aims to create a thinking, conscious machine that is equal or better than a human.
  - E.g. - perform any intellectual task that a human being can.

Strong AI as Computationalism or Functionalism

- In the context of the Chinese room argument, Searle has identified "strong AI" as "computer functionalism“.
- Functionalism holds that we can define mental phenomena (such as beliefs, desires, and perceptions) by describing their functions in relation to each other and to the outside world.
- Because a computer program can accurately represent functional relationships as relationships between symbols, a computer can have mental phenomena if it runs the right program.

Three tenets of computationalism

- The mind is a computer program.
  - Mental states are just implementations of computer program(s).
- The brain is irrelevant.
  - Mind ~ Computational state; Brain ~ Implementation
  - Computational states are implementation-independent.
- The Turing Test is decisive.
  - There is no stronger empirical test for the presence of mental states than Turing-Indistinguishability; hence the Turing Test is the decisive test for a computationalist theory of mental states.

Chinese Room Argument (CRA)

- Chinese Room Argument aims to refute the functionalist approach to understanding minds.
- The argument counts especially against that form of functionalism known as the Computational Theory of Mind that treats minds as information processing systems.
- In a nutshell, CRA refutes the possibility of strong AI.

What is Chinese Room Argument ?

- Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a **data base**) together with a book of instructions for manipulating the symbols (the **program**).

- The people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the **input**).

- By following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the **output**).

Chinese Room Argument…

- The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.

- Hence, if the man in the room does not understand Chinese on the basis of implementing the appropriate program for understanding Chinese then neither does any other digital computer solely on that basis because no computer has anything the man does not have."

Systems Reply : Objection

- The person in the room is part of a larger system. He doesn't understand Chinese, but the whole system (Man + room + rules) does understand Chinese.
- The knowledge of Chinese is in the rules contained in the room.
- The ability to implement that knowledge is in the man.
- Hence, the whole system understands Chinese.

Systems Reply : Rejoinder

- Let the individual internalize all of the system by memorizing the rules and script and doing the lookups and other operations in his head.

- Even now, he understands nothing of Chinese, and neither does the system, because there isn't anything in the system that isn't in him.

- If he doesn't understand then there is no way the system could understand because the system is just part of him.

Systems Reply : Rejoinder

- Further, it seems that if all we are requiring for intelligence is information processing, then everything can be seen as doing information processing. For e.g.
  - Stomach takes in something [food], processes it [digests it], and puts something out [energy]
  - But if our definition of intelligence is that it is 'information processing', why isn't the stomach intelligent?

Robot Reply : Objection

- What prevents the person in the Chinese room from attaching meanings to (and thus presents them from understanding) the Chinese ciphers is the sensory-motoric disconnection of the ciphers from the realities.

- Put a computer in the head of a robot giving the computer perceptual and motor capacities—this will bestow understanding and intentionality.

Robot Reply : Rejoinder

- Put the Chinese room in the robot's head
- All the robot's sensors do is provide additional input to the computer—and it will be just syntactic input.
- Searle argues that additional syntactic inputs will do nothing to allow the man to associate meanings with the Chinese characters.
- Hence, neither man nor does the robot understands Chinese.

Brain Simulator Reply : Objection

- Imagine a program that simulates the actual sequence of neuron firings at the synapses of a Chinese speaker when he understands stories in Chinese and gives answers to them.

- Surely then we would have to say that the machine understood the stories or else we would also have to deny that native Chinese speakers understood the stories.

Brain Simulator Reply : Rejoinder

- Instead of shuffling symbols, we have the man operate an elaborate set of water pipes with valves connecting them.

- Each valve corresponds to synapse in the Chinese brain, and turning on all the right valves, the Chinese answer pops out at the output end of the series of pipes.

- Searle claims that it is obvious that there would be no understanding of Chinese.

Brain Simulator Reply : Rejoinder

- As long as only the formal structure of the sequence of neuron firings at the synapses is simulated, it won't have simulated what matters about the brain, namely its causal properties, its ability to produce intentional states

- Intentionality is the property of being about something, having content.
  - Belief, desire are intentional states.

The Other Minds Reply : Objection

- We come to know that other people understand Chinese or anything else only by their behaviour.
- In principle a computer can also pass the behavioural tests.
- So if we are going to attribute cognition to other people we must in principle also attribute it to computers.

The Other Minds Reply : Rejoinder

- Searle argues that the problem in this discussion is not how I know that other people have cognitive states, but rather what it is that I am attributing to them when I attribute cognitive states to them.

- When we attribute cognitive states to other minds, it is more than complex behavioural dispositions. For Searle the additional is certain states of consciousness.

The Many Mansions Reply: Objection

- Searle's argument is helped along by the fact that present technology is limited.

- In future, other means besides programming might be devised such that computers may be imbued with whatever does suffice for intentionality by these other means.

The Many Mansions Reply: Rejoinder

- The original claim of strong AI is that mental processes are computational processes (programs) over formally defined elements.

- But this reply redefines the strong AI as whatever artificially produces and explains cognition.

- Hence, if strong AI is not identified with that precise, well defined thesis, CRA no longer apply because there is no longer a testable hypothesis for them to apply to.

Searle's Formal Derivation from Axioms

- (A1) Programs are formal (syntactic).
- (A2) Minds have mental contents (semantics).
- (A3) Syntax by itself is neither constitutive of nor sufficient for semantics.
- (A4) Brains cause minds.

- Conclusions
  - (C1) Programs are neither constitutive of nor sufficient for minds.
  - (C2) Any other system capable of causing minds would have to have causal powers (at least) equivalent to those of brains.

Further conclusions from Axioms

- (C3) Any artifact that produced mental phenomena, any artificial brain, would have to be able to duplicate the specific causal powers of brains, and it could not do that just by running a formal program.

- (C4) The way that human brains actually produce mental phenomena cannot be solely by virtue of running a computer program.

Conclusion

- It is important to recognize what Searle is not arguing for.
- He is not arguing that machines can't think.
- Searle is not arguing that thinking organisms necessary have to be made out of biological materials.
- It may be possible to produce a thinking machine made of non-biological materials. It is only that to do so, something more would be required than merely implementing a computer program.

Conclusion

- Finally, Searle is not arguing that thinking does not involve symbol manipulation. It is only that symbol manipulation cannot, by itself, constitute thinking.

References

- Searle, John. 1980a. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3, 417-424.

- Searle, J. (1990a), "Is the brain's mind a computer program?", *Scientific American* 262(1):26-31

- Harnad, S. (2001), “What's Wrong and Right About Searle's Chinese Room Argument?” *M. Bishop & J. Preston (eds.) Essays on Searle's Chinese Room Argument*. Oxford University Press.